5SRNA 二级结构的研究

1. 一种以Fox 模型为基础的微型计算机方法

刘次全 王 莹 周明培 李靖炎

摘要

本文系以 Fox 模型为兰本而设计的一种建立5S RNA二级结构的微型计算机方法。将5S RNA 分子的核苷酸顺序作为信息输入,经 Z—80微型计算机处理约 4 分钟后,即可建立在仅考虑二级相互作用的情况下,相当于 Fox 模型的具有最低自由能的二级结构。

本方法获得的结果与同样以 Fox 模型 为 兰本的其它方法所得到的结果相符。本方法简便、快速和费用低廉。易于推广应用。

迄今为止,有关 5SRNA 二级结构的研究,尽管运用了物理方法(如分析离心、紫外吸收、园二色谱、旋光色散、拉曼光谱、小角度 X 射线衍射和核磁共振等)、化学修饰法、部分酶解法和寡核苷酸结合法等等,但还没有得到一致公认的二级结构模型。研究表明,由于采用的研究方法不同,即使是同一种5S RNA分子,也会得出不同的二级结构图 (Erdman等,1976)。近几年来,在一些学者提出的模型中,Fox模型与大多数物理、化学方法测得的结果是相符的。就大肠杆菌5S RNA而言,该模型与核磁共振分析存在有28对碱基对的结果是一致的(李楠茜等,1982,Fox等,1975)。

为探讨 5SRNA二级结构的生物学意义,我们在 Fox 模型的基础上,设计了这种在 Z-80微型计算机上即可实现的建立5S RNA二级结构的计算机方法。同时,我们按照 自由能最低原则还提出了另一种模型(另文发表)。欲得到某种5S RNA 分子的类 Fox 模型的二级结构图,只需将已知的该分子的核苷酸顺序作为信息输入,经计算机处理后即可打印出所需的结果。 表 1 即是应用本方法所得到的大肠杆菌 (Esherichia Coli) 5SRNA 二级结构的计算机报告。

^{*}云南省电子计算站荣茂森同志参加了部分工作,特此致谢。

本文1983年3月10日枚到。

方 法

模型处理

我们是以 Fox 等 (Fox等, 1975) 提出的模型 (图 1 系大肠杆菌 5SRNA 二级结构 的Fox模型) 为兰本,按下述规则来处理的:

- 1.一个螺旋区是由连续的G一C、A-U和G-U破基对构成的。一个螺旋区至少需有三个碱基对。
- 2.碱基配对是: G跟C, G跟U和A跟U配对。然而, G—U碱基对不能在一个螺旋区的两端中的任何一端参与形成螺旋区(此点与Fox模型不同)。这是由于考虑到螺旋区的稳定性的缘故。
- - 图 1 大肠杆菌 (E. Coli) 5SRNA 二级结构模型 (Fox, 1975)

3.一个"发夹"突环必需包含至少 是三个碱基,其自由能则取决于关闭此"发夹"突环的碱基对(G—C或是A—U)。 下图表示由G—C对关闭的"发夹"突环:

4.一个"内突环"的自由能取决于究竟是由两对G-C对,或是一对G-C和一对 A-U对,还是两对A-U对关闭这一突环。下图表示由一对G-C和一对 A-U对关 的闭"内突环":

5.一个"突起"突环的自由能由横跨这一"突起"的两对碱基对的堆积自由能来确定。必须指出,在大肠杆菌 5SRNA 二级结构的 Fox 模型中没有"突起"突环。

螺旋区 (磺基对区) 是通过突环连接起来的。下图表示螺旋区 [和 [由 "突起"突环连接, [和 [之间由"内突环"连接。

自由能分布

从热力学知,一个稳定的 RNA 分子的二级结构应该使溶液的自由 能 最 低。 假 定 RNA分子的双链结构对单链结构的自由能记为 ΔG ,则:

$$\Delta G = \Delta H - T \Delta S$$

如果 Δ H和 Δ S 与绝对温度 T无关,于是RNA分子内双链形成的总自由能(Δ G_t)为碱基对起始形成的自由能(Δ G_i)与相继增长的自由能(Δ G_s)之和

$$\Delta G_{t} = \Delta G_{t} + \Delta G_{x}$$

$$= \Delta H \left(1 - \frac{T}{T_{x}}\right) - 2.3RT \log \gamma_{m}$$

式中Tin为熔解温度, Ym为m个未键合碱基的突环形成的概率, R为气体常数。 表 2 列出了本文采用的自由能数值* (Tinoco等, 1973)。

计算机程序

建立一种具有最低自由能的结构,能量函数是一关键因素。因为它可以排除不可能的或不需要的结构。必须强调指出,用一种数学处理来选择具有最低自由能的结构,其处理的量(结构数量)是非常庞大的,即使是 5SRNA 这样的小分子也是如此。

计算机程序系运用 FORTRAN N算法语言编制的两个子程序 组合 成的。表 1 所列自由能数值编制在计算机程序中。在输入给定的 5SRNA 分子的已知核苷 酸 顺序,如:

后,程序按G与C(或U)、A与U配对在螺旋区形成氢键,不能配对的碱基形成突环的原则,对所有可能的螺旋区进行编辑。

螺旋区由三对或三对以上碱基对顺次排列成反平行的键,如:

^{*}自由能数值按文献(Tinoco等, 1973) 作了调整和补充。

表 2

25°C RNA二级结构的自由能分布

碳 基 对 区 城	△G (千卡) ±10%
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	-1.2
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	- 1.8
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	- 2.2
-c-g- -g-c- -c-g-	-3.2
-c-c- -c-c- -c-c-	-5.0
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	-0.3
$G \longrightarrow G \longrightarrow$	-1.3 (內部的)
-G-G- -U-C-, -C-U-	-2.1 (末端的)
$\xrightarrow{-G-C-} \xrightarrow{-C-G-} \xrightarrow{-U-G-} \xrightarrow{-G-U-}$	-1.3 (內部的)
	~ 2.1 (末端的)
未结合区域的未结合碳基數	△G (千卡) ±1千卡
ALCOH M. WOLDSHA CLANE DE ANTONIO	"内突环"
2 — 6	+ 2
7-20	+ 3
m (>20)	$+1+2\log m$
	"突起" 突环
1	+ 3
2 — 3	+ 4
4 — 7	+ 5
8-20	+ 6
m (>20)	+ 4 + 2 log m
	"发夹突环"
	G一C美闭 A一U美闭
3	+ 8 > 8
4 — 5	+ 5 + 7
6 — 7	+ 4 + 6
8 9	+ 5 + 7
10-30	+ 6 + 8
m (>30)	3.2+2 log m 5.5+2 log m

在程序中每个 5S RNA 序列的碱基由顺序编号来表示: 5′端的第一个碱基编为 1号,依次为 2,3,4,5,……,直至3′端的第一个碱基(即序列的最后一个碱基)。在编辑时程序作如下处理:

5'端依次3个碱基与3'端的3个碱基相比较,

找出符合规则 1 和 2 的区域,然后将两端相匹配的序号分别存入MX,NX两个矩阵中,进而再把这些区域与其相应的自由能联系起来。最后,将重复的区域以及在能量上和空间上不利的区域除去, 从而找出以 Fox 模 型 为基础的,在仅考虑二级相互作用的情况下具有最低总自由能的结构。

本方法除去了螺旋区两端G一U碱基对对于螺旋区的自由能贡献(见规则 2),并注意到在相同碱基重复存在的地方对螺旋结构稳定性的影响。原则上根据螺旋形堆积的负自由能贡献和突环形成的正自由能贡献之和,就能直接计算以某种方式排列的5SRNA二级结构的稳定性。

结 果

根据我们对49个已知其核苷酸顺序的5SRNA 分子和酵母丙氨酸 t RNA分子二级结构的处理结果表明:

- 1) 用本文报告的方法建立相应于 Fox 模型的 5SRNA 和 "三叶草"型的 t RNA 分子可能的二级结构图是相当理想的。建立的二级结构跟有关文献报告的结果基本相符 (Darlix等, 1981, Delihas等, 1981, Fox等, 1975, Jordan等, 1974, Luchrsen等, 1981, Studnicka等, 1981)。
- 2) 本方法对小分子RNA 是普遍适用的,可用于任何直到140个碱基长度的多核苷酸链(对多于140个碱基长度的RNA分子,我们尚未试算过)。
- 3) 对每一种可能的二级结构同时给出一个总自由能数值,但处理的结果只打印出 具有最低总自由能值的二级结构。
- 4) 在一般微型计算机上即可进行。本方法就是在 Z -- 80 微型计算机上建立和通过的。处理一个5 SrRNA 分子的二级结构仅需 3 ~ 4 分钟,加上人工画出二级结构图的时间总共也不过 8 ~ 10 分钟。

本方法尤其具有简便、快速和费用低廉等特点,且易于推广应用。

讨 论

与其它工作的比较 总的来说,用本方法建立以Fox模型为基础的 5SRNA 的二级结构与同样以 Fox 模型为兰本的其它方法所得到的结果相符。

如前所述,就 5SRNA 而言,至今还没有一致公认的二级结构模型。因此,要建立一个能够指导合理选择二级结构模型的方法是困难的。本文只是在 Fox 关于大肠杆菌 5SRNA 二级结构模型的基础上,按照最低自由能来选择和建立 5SRNA 二级结构的。

由表 1 我们得到图 2。从图 1 和图 2 的对比中可以看出,两者的第 I 和第 I 螺旋区 完全相同,但第 IV螺旋区在图 2 较之图 I 多了三对碱基对 (即 | | |)。在 第 二 螺 旋

GCGGUG GCGC 区,图1为||||,图2为|||。图1 (Fox 模型) 的总能量是-47.30 千 UGCCGC CGCG

卡 (稳定性数是39),图 2 的总能量是 - 48.64千卡 (稳定性数是41)。

上述图 1 和图 2 产生部分差异的原因之一是:我们除去了螺旋区两端 G 一 U 碱基对的 贡献,这与 JAMES 等 (James等,1975)人的处理相同。

局限性 以 Fox 模型为 兰本和以自由能作判据本身就带有局限性,加之我们的人为处理,因而计算出的具有最低自由能的二级结构图并不一定符合真实的二级结构。此外,给出的自由能可可能是不够精确的。然而,在有多种可能的二级结构同时存在的情况下,判别具有最低自由能的首选结构无疑是十分重要的。必须强调指出,按照 Fox 模

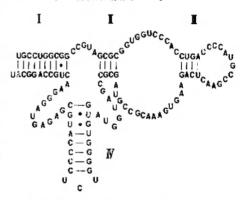


图 2 大肠杆菌 (E. Coli) 5SRNA 二級结构图

型,本方法规定任何两个螺旋区之间由"内突环"连接,而不包括如图所示的一个未配对减基的"突起"突环,

$$-C-G-G-C$$
 $G-A-A-C$
 $-G-U-C-G-C$
 $-C-U-U-C$

例如,如果出现下述情形,本方法将选取A而捨去B:

本文所述之方法在建立以Fox 模型为基础的 5SRNA 可能的二级结构时,是有一定的实际意义和应用价值的。

参考文献

李楠茜、赵翼、魏西平 1982 5SRNA 的结构,生物化学与生物物理进展(6):11-15。

Darlix, J. L. & Rochaix, J. D. 1981 Nucleotide Sequence and Structure of Cytoplasmic 5S RNA and 5.8S RNA of Chiamydomonas reinhardii. Nucl Acids Res. 3: 1291-1299.

Delihas, N., Andersen, J., Sprouse, H.M. & Dudoch, B.1981 The Nucleotide Sequence of Chloroplast 5S Ribosomal RNA from Spinach. Nucl. Acids Res. 9: 2801-2805.

Erdman, V. A. 1976 Structure and function of 5S and 5.8S RNA. In: Progress in Nucleic acid research and molecular biology, Vol. 18, Academic Press, New York, San Francisco, London 45-90.

Fox, G. E. & Woese, C. K. 1975 5S RNA Secondary structure. Nature 258: 505-507.

James, M. P. & James, E.M. 1975 Method for predicting RNA secondary structure. Proc. Nat. Acad. Sci. USA 72:2017-2021.

Jordan, B.R., Galling, G. & Jourdan, R. 1974 Sequence and Conformation of 5S RNA from Chlorella Cytoplasmic Ribosomes: Comparison with other 5S RNA Molecules. J. Mol. Biol. 87: 205-225.

Luehrsen, K. R., Fox, G.E., Kilpatrick, M. W., Walker, R.T., Domadey, J., Krupp, G & Gross, H. J. 1981 The nucleotide Sequence of the 5S rRNA from the archaebacterium Thermoplasma acidophilum. Nucl. Acids Res. 9: 965-970.

Studnicka, G. M., Eiserling, F. A. & Lake, J. A. 1981 A unique secondary folding pattern for 5S RNA corresponds to the lowest energy holologous secondary structure in 17 different prokaryotes. Nucl. Acids Res. 9: 1885—1904.

Tinoco, I., Borer, P. N., Dengler, B. & Levine, M. D. 1973 Improved estimation of secondary structure in Ribonucleic acids. Nature-New Biology 248: 40-41.

STUDIES OF 5SRNA SECONDARY STRUCTURE

1. A microcomputer method on the basis of the Fox model

Liu Ciquan, Wang Ying, Zhou Mingpei and Li Jingyan (Kunming Institute of Zoology, Academia Statica)

This microcomputer method is designed with the Fox model as reference for building 5SrRNA secondary structure. By inputing the nucleotide sequences of 5SrRNA and calculating through the work of the Z-80 microcomputer about 4 minutes, we got the secondary structure with the lowest total free energy, corresponding to the Fox model when only secondary interactions are under consideration. The result is basically in agreement with the literature (2, 4, 5, 6, 7, 8).

The method has the advantage of simplicity, low cost and comsuming less time, so, it is a method that can be easily popularized.

表 1 UGCCU GGCGG CCGUA GCGCG GUGGU CCCAC CUGAC CCCAU GCCGA ACUCA GAAGU GAAAC GCCGU AGCGC CGAUG GUAGU GUGGG GUCUC CCCAU GCGAG AGUAG GGAAC UGCCA GGCAU

NO.	NO. ENERGY				BAS	E															
2	-4.3	1	2	3	104	103	102														
3	-4.3	1	2	3	78	77	76														
4	- 4.3	1	2	3	66	65	64														
6	-9.3	1	2	3	4	78	77	76	75												
11	- 30.8	1	2	3	4	5	6	7	8	9	10	119	118	117	116	115	114	113	112	111 1	110
64	-9.2	11	12	13	14	76	75	74	73												
65	- 9.5	11	12	13	14	15	84	83	82	81	80										
66	-4.2	12	13	14	96	95	94														
67	- 4.2	12	13	14	75	74	73														
68	-5.4	12	13	14	61	60	59														
69	- 4.2	12	13	14	41	40	39														
70	-7.4	12	13	14	15	98	97	96	95												
71	-4.5	12	13	14	15	83	82	81	80												
73	- 4.2	13	14	15	97	96	95														
74	-4.2	13	14	15	42	41	40														
77	-8.2	16	17	18	62	61	60														
78	- 13.2	16	17	18	19	70	69	68	67												
79	-8.2	17	18	19	88	97	96														
80	-4.2	17	18	19	83	82	81														
81	-4.2	17	18	19	81	80	79														
82	-8.2	17	18	19	69	68	67														
63	-4.2	17	18	19		55	54														
84	- 8.2	18	19	20		69	68														
85	-8.2	18	19	20			60														
89	-8.2	19	20	21			70														
91	-8.2	19	20	21			42														
92	-4.2	19	20	21			31	71	70	69	68										
94	- 12.4	19	20	21			72 64	71 63		61											
95	-12.4	19	20	21			34	100	U2	01	00										
96	-7.2	20	21	22	36 31	35 30	29														
97	- 7.2 - 9.2	20	21	22			70	69	68												
98 99	- 9.2	20	21	22			62	61	60												
H 2	- 7.4			40			~														

-7.2

171

37 38 39 76 75 74 -9.0 37 38 39 40 76 75 74 73

```
-9.4 20 21 22 23 31 30 29 28
102 - 14.4
           20 21 22 23 24 31 30 29 28 27
    -4.2
           21 22 23 70 69 68
     -4.2
           21 22 23 62 61 60
     -4.4
186
           21 22 23 30 29 28
     -9.4
108
           21 22 23 24 30 29 28 27
     -7.2
111
           22 23 24 94 93 92
    -7.2
112
           22 23 24 39 38 37
     -7.2
           22 23 24 29 28 27
     -7.2
           23 24 25 36 35 34
    - 7.2
115
           23 24 25 31 30 29
     - 9.4
           23 24 25 26 36 35 34 33
           23 24 25 26 27 28 29 30 88 87 86 85 84 83 82 81
121 -18.9
    -4.4
           24 25 26 35 34 33
125 - 12.2
           25 26 27 28 108 107 106 105
126 -10.0
           26 27 28 107 106 105
130 -14.4
           26 27 28 29 30 85 84 83 82 81
    -7.2
132
           27 28 29 76 75 74
    -9.4
133
           27 28 29 30 84 83 82 81
           28 29
                 30 83 82 81
     -4.4
135
           28 29 30 81 80 79
    -4.4
           28 29 30 56 55 54
136
     -7.2
           29 30 31 87 86 85
138
     -7.2
           29 30 31 77 76 75
     -7.2
           30 31 32 106 105 104
     -6.6
           31 32 33 34 51 50 49 48
145
     -4.4
           32 33 34 50 49 48
151 - 26.2
           33 34 35 36 37 38 39 40 41 42 88 87 86 85 84 83 82 81 80 79
    -7.2
           34 35 36 77 76 75
           34 35 36 37 38 39 40 41 42 87 86 85 84 83 82 81 80 79
157 - 24.0
   - 10.0
           35 36 37 107 106 105
           35 36 37 38 39 40 41 42 86 85 84 83 82 81 80 79
163 - 21.8
164 - 10.0
           36 37 38 107 106 105
168 -16.8
           36 37 38 39 40 41 42 85 84 83 82 81 80 79
```

```
172 -11.8 37 38 39 40 41 42 84 83 82 81 80 79
```

173 -4.0 38 39 40 96 95 94

174 -4.0 38 39 40 75 74 73

175 -6.2 38 39 40 41 96 95 94 93

176 -6.8 38 39 40 41 42 83 82 81 80 79

178 -4.0 39 40 41 95 94 93

179 -4.3 39 40 41 65 64 63

181 -4.6 39 40 41 42 82 81 80 79

184 -4.3 40 41 42 104 103 102

186 -4.3 40 41 42 66 65 64

188 -9.3 40 41 42 43 78 77 76 75

190 -4.4 46 47 48 103 102 101

191 -4.4 46 47 48 80 79 78

192 -4.4 46 47 48 55 54 53

193 -6.6 46 47 48 49 103 102 101 100

194 -6.4 46 47 48 49 87 86 85 84

195 -4.2 47 48 49 107 106 105

196 -4.4 47 48 49 102 101 100

197 -4.4 47 48 49 100 99 98

200 -6.4 47 48 49 50 85 84 83 82

201 -4.4 48 49 50 57 56 55

203 -4,4 50 51 52 89 88 87

205 -4.2 54 55 56 70 69 68

206 -4.2 54 55 56 62 61 60

207 -5.4 59 60 61 65 64 63

208 -6.4 59 60 61 62 82 81 80 79

209 -8.2 60 61 62 98 97 96

210 -4.2 60 61 62 83 82 81

211 -4.2 60 61 62 81 80 79

212 -8.2 60 61 62 69 68 67

214 -9.2 62 63 64 65 76 75 74 73

215 -9.5 62 63 64 65 66 84 83 82 81 bi

216 -4.2 63 64 65 96 95 94

217 -4.2 63 64 65 75 74 73

218 -7.4 63 64 65 66 98 97 96 95

219 -4.5 63 64 65 66 83 82 81 80

```
221 -4.2 64 65 66 797 96 95
   -8.2 68 69 70 98 97 96
223
   -4.2 68 69 70 81 80 79
926
   -9.0 73 74 75 76 95 94 93 92
   -7.2 74 75 76 94 93 92
239
   -4.2 76 77 78 97 96 95
247 -21.7 79 80 83 82 83 84 85 86 97 96 95 94 93 92 91 90
252 -17.2 82 83 84 85 86 94 93 92 91 90
253 -10.0 83 84 85 93 92 91
254 -10.0 83 84 85 92 91 90
256 -15.0 83 84 85 86 93 92 91 90
   -9.2 83 84 85 86 91 90 89 88
258 -10.0 84 85 86 93 92 91
   -10.0 84 85 86 92 91 90
   -4.2 84 85 86 90 89 88
260
   -6.4 87 88 89 90 108 107 106 105
264
   -6.6 87 88 89 90 101 100 99 98
   -4.2 88 89 90 107 106 105
268 -4.4 88 89 90 100 99 98
270 -12.2 89 90 91 92 108 107 106 105
271 -10.0 90 91 92 107 106 105
  11 78 144 247
BASE PAIR 1 BOND 119
                        UA
                             ENERGY =
                             ENERGY = -2.200 TOTAL -2.20
                        GC
BASE PAIR 2 BONC 118
                             ENERGY = -5.000 TOTAL -7.20
                        CG
BASE PAIR 3 BOND 117
                             ENERGY = -5.000 TOTAL -12.20
BASE PAIR 4 BOND 116
                        CG
                            ENERGY = - 2.200 TOTAL -14.40
BASE PAIR 5 BOND 115
                        UA
                             ENERGY = -2.200 TOTAL -18.60
BASE PAIR & BOND 114
                        GC
                             ENERGY = -5.000 TOTAL -21.60
BASE PAIR 7 BOND 113
                        GC
                             ENERGY = -5.000 TOTAL -26.60
BASE PAIR 8 BOND 112
                        CG
                             ENERGY = -2.100 TOTAL -28.70
                        GU
BASE PAIR 9 BOND 111
                             ENERGY = -2.100 TOTAL -30.80
BASE PAIR 10 BOND 110
                        GC
                             3.8
INITIAL RING
                             ENERGY =
                        GC
BASE PAIR 16 BOND 70
                         CG ENERGY = - 5.000 TOTAL - 5.00
```

BASE PAIR 17 BOND 69

BASE	PAIR	18	BOND	68	GC	ENERGY = -3.200	TOTAL - 8.20
BASE	PAIR	19	BOND	67	CG	ENERGY = -5.000	TOTAL - 13.20
INITIA	L RIN	G				3.8	
BASE	PAIR	31	BOND	51	CG	ENERGY =	
Ditota	11111	•					
BASE	PAIR	32	BOND	50	UA	ENERGY = -2.200	TOTAL - 2.20
BASE	PAIR	33	BOND	49	GC	ENERGY = -2.200	TOTAL - 4.40
BASE	PAIR	34	BOND	48	AU	ENERGY = -2.200	TOTAL - 6.60
HAIR		35	TO	47		8.0	
BASE	PAIR	79	BOND	97	GC	ENERGY =	
BASE	PAIR	80	BOND	96	UG	ENERGY = -2.100	TOTAL - 2.10
BASE	PAIR	81	BOND	95	GU	ENERGY =300	TOTAL - 2.40
BASE	PAIR	82	BOND	94	UA	ENERGY = -2.100	TOTAL - 4.50
BASE	PAIR	83	BOND	93	GC	ENERGY = -2.200	TOTAL - 6.70
BASE	PAIR	84	BOND	92	GC	ENERGY = -5.000	TOTAL - 11.70
BASE	PAIR	85	BOND	91	GC	ENERGY = -5.000	TOTAL - 16.70
BASE	PAIR	86	BOND	90	GC	ENERGY = -5.000	TOTAL - 21.70
HAIR		87	TO	89		8.0	

TOTAL ENERGEN SUM = -48.64